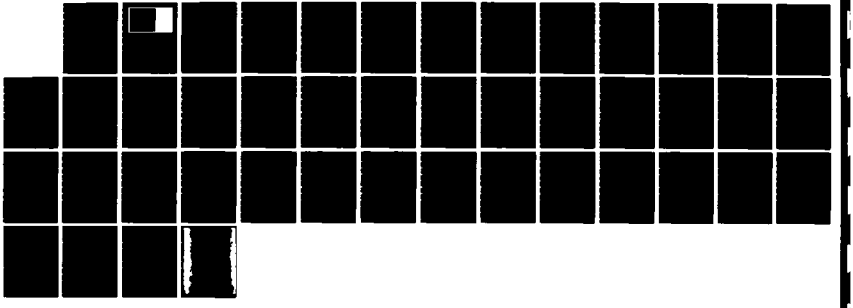
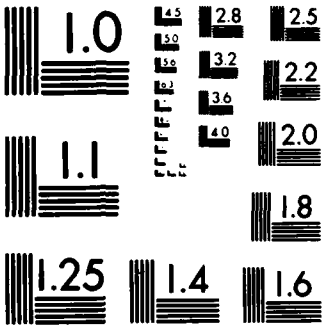
AD-A142 843 BAYESIAN MODELS FOR RESPONSE SURFACES I THE EQUIVALENCE OF SEVERAL MODELS..(U) WISCONSIN UNIV-MADISON MATHEMATICS RESEARCH CENTER D M STEINBERG APR 84 1/1

UNCLASSIFIED MRC-TSR-2682 DAAG29-80-C-0041 F/G 12/1 NL





MRC Technical Summary Report #2682

BAYESIAN MODELS FOR RESPONSE SURFACES I: THE EQUIVALENCE OF SEVERAL MODELS AND THEIR RELATIONSHIP TO SMOOTHING SPLINES

David M. Steinberg

Mathematics Research Center
University of Wisconsin—Madison
610 Walnut Street
Madison, Wisconsin 53705

April 1984

(Received April 16, 1984)

Approved for public release
Distribution unlimited

Sponsored by

U. S. Army Research Office
P. O. Box 12211
Research Triangle Park
North Carolina 27709

National Science Foundation
Washington, D. C. 20550

UNIVERSITY OF WISCONSIN-MADISON
MATHEMATICS RESEARCH CENTER

BAYESIAN MODELS FOR RESPONSE SURFACES I: THE FQUIVALENCE OF SEVERAL MODELS AND THEIR RELATIONSHIP TO SMOOTHING SPLINES*

David M. Steinberg

Technical Summary Report #2682
April 1984

ABSTRACT

In response surface modeling, simple graduating functions such as low-degree polynomials are used to approximate complex, unknown response functions. Several authors have suggested Bayesian generalizations of response surface models that incorporate prior belief as to the (in)adequacy of a graduating function to represent a response function. We show that the models of Smith (1 73), Blight and Ott (1975), and O'Hagan (1978) are equivalent statements. We also show how their models are related to the generalized smoothing splines of Wahba (1978) and to Young's (1977) proposal for Bayesian polynomial regression. Finally, we suggest a canonical representation of the models in terms of generalized Fourier series expansions of the response function and show how such expansions can be used to develop reasonable prior distributions.

AMS (MOS) Subject Classifications: 62F15, 62J05

Key Words: Response Surface Models; Bayesian Linear Model; Hierarchical Linear Model; Localized Regression Model; Smoothing Splines; Polynomial Regression; Series Estimators

Work Unit Number 4 (Statistics and Probability)

---

*This work forms part of the author's doctoral dissertation written under the direction of Professor G. E. P. Box. The author is grateful to Professor Box for his many valuable comments.

This research was sponsored by the United States Army under Contract No. DAAG29-80-C-0041. This material is based upon work supported by the National Science Foundation under Grant No. MCS-8210950.

SIGNIFICANCE AND EXPLANATION

Scientists often wish to describe the relationship between a response variable and a collection of explanatory variables. When the particular nature of the relationship is unknown, as is often the case, a common strategy is to develop an empirical model by using a simple graduating function such as a low-degree polynomial to approximate the true relationship. The techniques of response surface methodology were developed to accomplish this goal.

Several authors have proposed generalizations of standard response surface models that attempt to take into account the approximate nature of the graduating functions that are used. In this paper we show that the models of Smith (1973), Blight and Ott (1975), and O'Hagan (1978) are equivalent to one another. These models share a common Bayesian approach in which probability distributions are used to reflect the scientist's prior beliefs about the (in)adequacy of the graduating function to represent the true response function. We also show how the models are related to Wahba's (1978) generalized smoothing splines and to Young's (1977) Bayesian approach to polynomial regression. Finally, we consider a Bayesian model that involves an expansion of the response function as a convergent series of functions, with special attention to an expansion using Hermite polynomials.

BAYESIAN MODELS FOR RESPONSE SURFACES I: THE EQUIVALENCE OF SEVERAL MODELS AND THEIR RELATIONSHIP TO SMOOTHING SPLINES*

David M. Steinberg

1. INTRODUCTION

Many scientific investigations are designed to explore the relationship between a response variable Y and a set of explanatory variables, $X_1,\ldots,X_k$. Sometimes the physical nature of the problem suggests a specific functional form linking the response to the input variables. Often, however, the functional nature of the response is either unknown or is too complicated to provide a useful representation. A strategy that is often employed in these situations is to seek an empirical model which, it is hoped, will provide a good local approximation to the response function for those combinations of the explanatory variables considered to be of greatest interest.

An important body of statistical techniques that has been developed for empirical modeling problems in which all or most of the explanatory variables are continuous is known as response surface methodology (see, for example, Box and Wilson 1951, Box 1954, Box and Youle 1955, Myers 1976). Traditionally, response surface models have exploited simple graduating functions, such as low-degree polynomials, to approximate the true response function. Section 2 describes these models and establishes some notation.

Several authors have proposed Bayesian generalizations of classical response surface models that are designed to take into account the approximate nature of empirical graduating functions. Section 3 discusses the rationale behind the Bayesian approach to response surface models, compares it to

---

*This work forms part of the author's doctoral dissertation written under the direction of Professor G. E. P. Box. The author is grateful to Professor Box for his many valuable comments.

This research was sponsored by the United States Army under Contract No. DAAG29-80-C-0041. This material is based upon work supported by the National Science Foundation under Grant No. MCS-8210950.

similar Bayesian models for other estimation problems, and then discusses the models suggested by Smith (1973), Blight and Ott (1975), and O'Hagan (1978). In particular, we show that these three models, although expressed in different forms and justified by different arguments, are in fact equivalent statements. Section 4 shows how these Bayesian models are related to the generalized smoothing splines of Wahba (1978). Section 5 describes a canonical form for the models in terms of generalized Fourier series expansions of the response function, discusses the significance of assuming an improper prior for the regression coefficients, and indicates how the models are related to Young's (1977) Bayesian method for polynomial regression and to ridge regression. Section 6 considers a particular application of the generalized Fourier series approach to develop a *reasonable prior* distribution, and Section 7 summarizes the results and discusses the use of Bayesian models to represent model inadequacy.

The implications of the Bayesian models for estimating a response surface will be described in a sequel to this paper.

2 CLASSICAL RESPONSE SURFACE MODELS

Response surface models were first proposed by Box and Wilson (1951) as a technique to study the relationship between an observed experimental response variable Y and a set of continuous explanatory variables $X_1,\ldots,X_k$, with the goal of finding settings of the explanatory variables that optimize the response. The explanatory variables might be the raw inputs to the system or

suitably transformed functions of the raw inputs (found, say, by transforming to a more appropriate metric or simply by centering and scaling).

Suppose the true response function relating the response variable to the explanatory variables is $g(x)$, where x denotes a point in the explanatory variable space. The basic idea behind response surface models is to approximate g by a simple graduating function, at least over a limited region of interest in the explanatory variable space. The graduating functions which have been used most often in response surface models are low-degree polynomials, and the simplest of these is a first degree polynomial:

$$g(x) \approx \beta_0 + \sum_{i=1}^{k} \beta_i x_i. \qquad (2.1)$$

If (2.1) is judged to be an inadequate representation of the true response function, a second degree polynomial might be used:

$$g(x) \approx \beta_0 + \sum_{i=1}^{k} \beta_i x_i + \sum_{i=1}^{k} \sum_{j=i}^{k} \beta_{ij} x_i x_j. \qquad (2.2)$$

Polynomial models of higher degree can be defined in an analogous manner, with the d'th degree polynomial including all terms of the form:

$$\prod_{i=1}^{k} x_i^{\alpha(i)},$$

where the $\alpha(i)$ are non-negative integers whose sum is less than or equal to d.

In order to estimate the unknown parameters in a polynomial model, experimental data $\{Y_i, x_i\}_{i=1}^{n}$ must be gathered in which the response variable is observed at n settings of the explanatory variables. For any polynomial graduating function, the ith data point can be modeled as:

$$Y_i \approx f(x_i)'\beta + \varepsilon_i, \tag{2.3}$$

where f is a vector of functions whose elements are the appropriate powers of x_i, β is a vector of coefficients that must be estimated from the data, ε_i denotes experimental error, and primes denote transposes. The entire data vector Y can then be written as the approximate linear model:

$$Y \approx X\beta + \varepsilon, \tag{2.4}$$

where X is the matrix whose ith row is $f(x_i)'$.

The polynomial models defined above are useful when the explanatory variables are continuous, but special consideration is necessary for categorical variables. The true response function will not be continuous with respect to categorical input variables so that it does not make sense to attempt to "graduate" the response between levels of a categorical variable. Box (1954) suggested that the best approach in an experiment involving both continuous and categorical inputs would be to carry out a separate investigation at each categorical factor combination. Such terms can be included in the general linear model (2.4) simply by adding appropriate elements

to f. Thus (2.4) is also appropriate when there are both continuous and categorical explanatory variables.

3. BAYESIAN RESPONSE SURFACE MODELS

The regression function in a response surface model is chosen with the hope that it will provide a good local approximation to the true response function. Analysis of the model typically proceeds, however, as though the regression function were an exact representation. The stimulus for the Bayesian models that will be discussed here is an attempt to achieve a more realistic model by describing the uncertainty about the response function in terms of prior probability distributions. This section will describe the models that have been proposed by Smith (1973), Blight and Ott (1975), and O'Hagan (1978). Each employs prior distributions to reflect the extent to which an empirical graduating function is believed to provide an adequate approximation to the true response function and, although each does so in a different way, we will show that the three models are in fact equivalent.

3.1 Smith's Hierarchical Model

Smith (1973) proposed a hierarchical Bayesian linear model to represent the relationship between a response vector $\mathbf{Y}$ and a matrix $\mathbf{X}$ of regressor variables associated with the dose administered in a dose-response experiment. It is simple and straightforward to extend Smith's model to arbitrary response

surface models and we do so here. The hierarchical structure consists of three tiers and provides the mechanism for building prior uncertainty about the response function into the statistical model. The model is a special case of the general three-tiered Bayesian linear model analyzed by Lindley and Smith (1972) so that all of their results may be applied here.

The model reads as follows:

$$Y/\theta_1 \sim N(\theta_1, \sigma^2 I). \tag{3.1a}$$

$$\theta_1/\theta_2 \sim N(X\theta_2, V). \tag{3.1b}$$

$$\theta_2 \sim N(\theta_3, V_1). \tag{3.1c}$$

The first tier of the model (3.1a) simply states that the observed responses Y are normally distributed and vary about their respective expected values θ_1 with common variance σ^2; the assumption of normality here is exactly analogous to the common assumption in linear model theory of normally distributed error terms.

The second tier (3.1b) invokes the linear model structure by asserting that the vector of expected values, θ_1, has a multivariate normal distribution with mean vector $X\theta_2$, where θ_2 is a vector of regression coefficients and corresponds directly to β in equation (2.4). The variance matrix V indicates the experimenter's a priori confidence in the adequacy of the linear model. If the elements of V are all quite small, then the model claims that the expected value vector θ_1 follows the linear model $X\theta_2$ closely; i.e., the linear model is assumed to be a good

representation of the true response function. If, on the other hand, the elements of V are rather large, this reflects prior belief that the true response may deviate considerably from the linear model, even though it may be the best current guess for the response function.

The final tier of the model simply assigns a prior distribution to the regression parameters. A diffuse prior is often deemed appropriate for the regression parameters and, following the argument of Lindley and Smith (1972), this can be achieved by considering limiting forms as V_1^{-1} converges to 0.

There is an interesting difference between Smith's model and most hierarchical Bayesian models. Such models are usually formed by taking as the first tier a conventional parametric sampling theory model that depends on some unknown parameters. Each successive tier of the model states a prior distribution for the (hyper)parameters appearing in the previous tier. Smith's model differs from this approach in that the hierarchical structure is used to break up the conventional linear model $Y = X\beta + \varepsilon$ into two different tiers, with the mediating parameter θ_1 separating the observed response vector from the linear model. We find this approach intriguing and wonder if it might not be useful in other contexts, as well.

3.2 Blight and Ott's Approximating Function + Bias Model

Blight and Ott (1975) proposed a Bayesian model for polynomial

regression. They considered only experiments with a single explanatory variable, but their ideas, like Smith's, can easily be extended to handle more general response surface problems and the presentation here will be appropriate for any number of explanatory variables. Their model represents each experimental response as a sum of three components:

$$\begin{aligned} \text{Response} &= \text{Low-degree polynomial approximation} \\ &+ \text{deterministic error (bias)} \\ &+ \text{random (experimental) error.} \end{aligned} \qquad (3.2)$$

The first term is a classical response surface model such as (2.3) and the last term is identical to the random error term in (2.3). What distinguishes Blight and Ott's model is the second term, which is an explicit statement of the approximate nature of the polynomial.

Mathematically, Blight and Ott's model for the ith observation can be written:

$$Y_i = f(x_i)'\beta + \eta_i + \varepsilon_i. \qquad (3.3)$$

The three terms on the right-hand side of (3.3) correspond to the respective components of (3.2). Observe that (3.3) is identical to the classical response surface model (2.3) but for the addition of the "bias" term, η_i, and the assumption that this term permits an exact representation of Y_i, so that an equals sign is now justified.

Blight and Ott completed their model specification by making the following distributional assumptions:

$$\beta \sim N(\beta_0, V_1). \tag{3.4a}$$

$$\eta \sim N(0, V), \text{ where } \eta' = (\eta_1, \ldots, \eta_n). \tag{3.4b}$$

$$\varepsilon \sim N(0, \sigma^2 I), \text{ where } \varepsilon' = (\varepsilon_1, \ldots, \varepsilon_n). \tag{3.4c}$$

η and ε are distributed independently. (3.4d)

Equation (3.4b) is more general than the assumption actually made by Blight and Ott, who stated a specific form for the elements of the matrix V that they felt would be appropriate for the polynomial regression situation studied in their paper.

A simple rationale underlies the distributional assumptions. Equation (3.4a) provides a prior distribution for the regression parameters and is directly analogous to (3.1c) in Smith's model. Again, a diffuse prior for the regression coefficients can be entertained by considering limiting forms as V_1^{-1} tends to a 0 matrix. Assumption (3.4c) is identical to that in the standard linear model.

The distribution of the vector η of bias terms given in (3.4b) is justified by appealing to prior belief about the ability of the linear response surface model to represent the true response function. The bias term represents that part of the response function not captured by the approximating polynomial. Since the approximating polynomial typically represents the best current guess as to how the response depends on the explanatory variables, it is reasonable to assign the bias at any point a prior mean of 0. The variance matrix in (3.4b) should suggest the possible severity of the bias. The diagonal elements of V can be interpreted as

reflecting the suspected magnitude of the bias at the respective design points; the off-diagonal elements reflect prior assumptions about how similar the bias is likely to be at corresponding pairs of design points which are closely related to prior convictions about the smoothness of the response function, since a response function which is smooth will have similar biases at proximate design points.

Theorem 3.1 The Smith model and the Blight-Ott model are mathematically equivalent.

Proof: The proof is quite simple and relies on a trivial re-writing of Smith's model. We simply write each of the first two stages in Smith's model as the sum of a deterministic term (the expected value) plus a random term with an appropriate covariance matrix. Thus, we rewrite equation (3.1a) as:

$$\mathbf{Y} = \theta_1 + \varepsilon, \quad \text{where } \varepsilon \sim N(\mathbf{0}, \sigma^2 \mathbf{I}). \tag{3.5a}$$

Similarly, we rewrite equation (3.1b) as:

$$\theta_1 = \mathbf{X}\theta_2 + \eta, \quad \text{where } \eta \sim N(\mathbf{0}, \mathbf{V}). \tag{3.5b}$$

Now, substituting (3.5b) into (3.5a) gives:

$$\mathbf{Y} = \mathbf{X}\theta_2 + \eta + \varepsilon, \tag{3.5c}$$

where the distributions of η and ε are given above, the distribution of θ_2 is given in (3.1c), and the three terms are independent. This is precisely the model for $\mathbf{Y}$ suggested by Blight and Ott, with θ_2 in place of $\boldsymbol{\beta}$.

3.3 O'Hagan's Localized Regression Model

O'Hagan (1978) suggested a different way to modify (2.3) to

reflect uncertainty as to the form of the response function. He argued that, while (2.3) may be adequate to describe the response function in the immediate neighborhood of any particular point $\mathbf{x} = (X_1,\ldots,X_k)$, it is unlikely to be valid over the entire range of explanatory variable settings that might be used. This led him to generalize (2.3) by allowing the parameter vector β to be a function of $\mathbf{x}$, characterizing the manner in which β varies with $\mathbf{x}$ in terms of a prior probability distribution. O'Hagan called this the "localized regression model."

O'Hagan formally defined the localized regression model by specifying the appropriate distributional assumptions for each point $\mathbf{x}$ in the explanatory variable space. Denoting by $Y_{\mathbf{x}}$ an observation at the point $\mathbf{x}$, he assumed that:

$$Y_{\mathbf{x}}/\beta(\mathbf{x}) \sim N(\mathbf{f}(\mathbf{x})'\beta(\mathbf{x}),\sigma^2), \tag{3.6a}$$

$$\beta(\mathbf{x})/\mathbf{b}_0 \sim N(\mathbf{b}_0,\mathbf{W}(\mathbf{x},\mathbf{x})). \tag{3.6b}$$

Finally, he assumed that the joint distribution of the $\beta(\mathbf{x})$ was normal with covariance function given by:

$$E\{[\beta(\mathbf{x}_1) - \mathbf{b}_0][\beta(\mathbf{x}_2) - \mathbf{b}_0]'/\mathbf{b}_0\} = \mathbf{W}(\mathbf{x}_1,\mathbf{x}_2). \tag{3.6c}$$

We can interpret $\mathbf{b}_0$ as the parameters of a global regression function about which there is local variation. When prior information does not suggest a specific global regression function, O'Hagan advocated using a vague prior distribution for $\mathbf{b}_0$, which can be accomplished by assuming that:

$$\mathbf{b}_0 \sim N(\mathbf{0},k\mathbf{I}),$$

and considering limiting forms as $k \to \infty$.

The matrix W in (3.6c) reflects the extent to which the parameters values are believed, a priori, to vary from one point to another. Thus W, like the matrix V in the Blight-Ott model, is related to prior beliefs about the smoothness of the response function. Large diagonal elements in W reflect prior belief that the parameters may fluctuate considerably, while large off-diagonal elements suggest that the parameter values should be quite similar at the respective points. O'Hagan was aware that the Blight-Ott model "shows many similarities" to his own (p. 23). However, concentrating on the specific covariance function analyzed in detail by Blight and Ott, he concluded that their model is a special case of the localized regression model. By considering Blight and Ott's model in the more general form described in (3.2) and (3.3), we now show that it is actually equivalent to O'Hagan's model.

Theorem 3.2: The model specification of (3.6a-c) is identical to that of (3.2)-(3.4), with b_0 in place of β and $V_{ij} = f(x_i)'W(x_i,x_j)f(x_j)$. The two models are equivalent if the vector of regression functions f includes a constant function.

Proof: The proof parallels that of Theorem 3.1. We begin by rewriting (3.6a) as:

$$Y_x = f(x)'\beta(x) + \varepsilon_x, \tag{3.7a}$$

where $\varepsilon_x \sim N(0,\sigma^2)$. Now rewrite (3.6b) as:

$$\beta(x) = b_0 + \zeta(x), \tag{3.7b}$$

where $\zeta(x) \sim N(0,W(x,x))$. Substituting (3.7b) into (3.7a):

$$Y_x = f(x)'b_0 + f(x)'\zeta(x) + \varepsilon_x$$

$$= f(x)'b_0 + \eta_x + \varepsilon_x, \quad \text{where} \quad \eta_x = f(x)'\zeta(x).$$

This is precisely the form of (3.3). All that remains to complete the proof is to show that $\eta' = (\eta_1,\ldots,\eta_n)$ has the distribution claimed in the theorem, and this can be trivially verified. Note that the reverse implication will be true if and only if the covariance function $V(x_1,x_2)$ for η_x speicified in a Blight-Ott model can be expressed in the form $f(x_1)'W(x_1,x_2)f(x_2)$ corresponding to a localized regression model. If the vector f includes a constant term (say the first element in f), then setting $W_{1,1}(x_1,x_2)$ equal to $V(x_1,x_2)$ and making the other entries in W equal to 0 reproduces the Blight-Ott model. If the vector f does not include a constant function, it is easy to show examples of covariance functions V that cannot be achieved by a localized regression model. It seems rather unlikely that a localized regression model would be used without a constant term, so this restriction is of no practical significance.

It should be noted that O'Hagan (1978) also proposed a generalization of the localized regression model that allowed for a vector-valued response variable, non-homogeneous error variances, and a general explanatory variable space (he restricted the localized model to a single explanatory variable). He also allowed for the prior expectation of the response variable to be an arbitrary function, not necessarily a polynomial of low degree. None of these generalizations affects the above Theorem. The extensions to a vector response and to non-homogeneous error

variance are straightforward and could be applied just as easily to Blight and Ott's model. The assumption of a general explanatory variable space has already been incorporated above. The prior expectation function was assumed to be a low-degree polynomial by Blight and Ott, but their model could also be used with any other type of approximating function, be it a fixed function or a parametric function with unknown parameters. Thus the correspondence between O'Hagan's model and Blight and Ott's model is valid also for O'Hagan's generalized model.

4. GENERALIZED SMOOTHING SPLINES

The spline function approach, on the surface, appears quite unrelated to the models described in Section 3. However, results of Wahba (1978) show that generalized smoothing splines are equivalent to the Bayesian response surface models when the regression coefficients are assigned a diffuse prior.

Generalized smoothing splines for estimating a response function of uncertain form were derived as solutions to a problem in functional approximation: find that member of a specified function space that most closely fits the observed data subject to a smoothness restraint. The solution in the general case exploits the structure of reproducing kernel Hilbert spaces (r.k.h.s.) (see Aronszajn (1950) for the general theory of r.k.h.s.). First, denote by $\{f_j\}_{j=1}^{p}$ the functions that constitute the elements of the vector f in (2.3). For standard response surface models, the

f_j will simply be the monomials that appear in the approximating polynomial, although in theory any desired collection of regression functions might be used. Let H_K be a r.k.h.s. of functions defined on the explanatory variable space that contains the f_j and has reproducing kernel $K(x_1,x_2)$. It can be shown that H_K has a representation as the direct sum of span $\{f_1,\ldots,f_p\}$ and a r.k.h.s. H_Q, which has reproducing kernel $Q(x_1,x_2)$. Let P_Q be the orthogonal projection operator from H_K onto H_Q. Then the generalized smoothing spline $g_{n,\lambda}$ is defined as the solution to the problem: find $g \in H_K$ to minimize

$$n^{-1}\sum_{i=1}^{n}[g(x_i) - Y_i]^2 \quad + \quad \lambda\|P_Q g\|_K^2 \qquad (4.1)$$

where the summation is over the n observed data points and the latter term is the squared norm (in H_K) of the projection of g onto H_Q times a smoothing parameter λ.

Much of the work on smoothing splines has focused on the case where the design space is the interval $[0,1]$, H_K is the Sobolev space: $W_2^p = \{g: g,g',\ldots,g^{(p-1)}$ abs. cont. $g^{(p)} \in L_2[0,1]\}$, $f_j(x) = x^{j-1}$, $j=1,\ldots,p$ and $\|P_Q(g)\|_K^2 = \int (d^p g/dx^p)^2 dx$. In this case, it is well known that $g_{n,\lambda}$ is a polynomial spline of degree $2p-1$ and is uniquely determined provided the data cannot be exactly interpolated by the approximating polynomial (see Wahba 1978). A common choice for p has been $p=2$, in which case $\|P_Q(g)\|^2 = \int (g^{(2)}(x))^2 dx$ and has a direct interpretation as a

measure of the smoothness of the solution. The choice of λ controls the tradeoff between how smooth the solution will be and how closely it will match the observed data.

Wahba (1978) proved the following theorem which relates spline smoothing to Bayesian estimation of a stochastic process.

Theorem 4.1: Suppose the true response function is $g(x)$, so that the ith data point is

$$Y_i = g(x_i) + \varepsilon_i,$$

where $\varepsilon = (\varepsilon_1,\ldots,\varepsilon_n)' \sim N(0,\sigma^2 I)$. Suppose the prior distribution of $g(x)$ is the same as that of the stochastic process

$$T_\xi(x) = \sum_{j=1}^{p} \beta_j f_j(x) + b^{1/2} Z(x), \qquad (4.2)$$

where $\beta = (\beta_1,\ldots,\beta_p)' \sim N(\beta_0,\xi I)$, $b>0$ is fixed and $Z(x)$ is a zero mean Gaussian stochastic process with $E\{Z(x_1)Z(x_2)\} = Q(x_1,x_2)$. Then for any fixed point x,

$$g_{n,\lambda}(x) = \lim_{\xi\to\infty} E_\xi\{g(x)/Y=y\},$$

where $\lambda = \sigma^2/nb$ and E_ξ denotes expectation with respect to the posterior distribution of $g(x)$ given the prior (4.2). Thus the smoothing spline solution $g_{n,\lambda}$ is the limiting posterior expectation of the response function given (4.2) when the prior distribution of the parameters in the approximating polynomial is made diffuse.

The characterization of spline smoothing in Theorem 4.1 as a form of Bayesian estimation suggests a similarity with the models defined in Section 3. We prove this in the following theorem.

Theorem 4.2: Under the prior specification (4.2) of the last theorem, the prior distribution of the data vector Y is given by the Blight-Ott model ((3.3) and (3.4)) with $V_1 = \xi I$ and with $V_{i,j} = bQ(x_i, x_j)$.

Proof: The i'th observation is $Y_i = g(x_i) + \varepsilon_i$. Then, given (4.2), the prior distribution for the i'th observation is the same as the distribution of

$$\sum_{j=1}^{p} \beta_j f_j(x_i) + b^{1/2} Z(x_i) + \varepsilon_i$$

$$= \sum_{j=1}^{p} \beta_j f_j(x_i) + \eta_i + \varepsilon_i.$$

The full data vector Y thus has a prior distribution identical to the distribution of

$$X\beta + \eta + \varepsilon$$

where X is an nxp matrix with $X_{i,j} = f_j(x_i)$, $\beta = (\beta_1, \ldots, \beta_p)'$, and $\eta = (\eta_1, \ldots, \eta_n)'$. The prior distributions of β, η, and ε are easily seen to be those claimed in the theorem.

Thus Theorems 4.1 and 4.2 demonstrate that the Bayesian models proposed by Smith, Blight and Ott, and O'Hagan all give rise to generalized spline estimates of the response function when the regression coefficients are assigned a vague prior distribution.

5. A GENERALIZED FOURIER SERIES APPROACH

The use of a generalized Fourier series to represent a response function is a broad generalization of the classical response surface approach. Whereas the classical response surface models described in Section 2 consist of a linear combination of a small number of simple graduating functions (e.g. the monomials which constitute a low degree polynomial), the generalized Fourier series models will provide an exact representation of the response function as a linear combination of an infinite sequence of functions. This section will show that the generalized Fourier series approach is equivalent to the Bayesian models described in Section 3 when appropriate prior assumptions are made about the coefficients in the series. The form of the generalized Fourier series models will then be exploited to discuss the significance of assigning the regression coefficients an improper prior and to point out relationships with ridge regression, multiple regression, and Young's (1977) Bayesian approach to polynomial regression.

5.1 Representing the Response Function

Denote the explanatory variable space by X, let μ be any σ-finite measure on X, and suppose the true response function $g(x)$ belongs to the space $L^2(\mu)$. It is well known that $L^2(\mu)$ is a separable Hilbert space (see Rudin 1974, p. 81), so that any function in the space can be represented in terms of a sequence of basis functions, much as a vector space can be decomposed using a

sequence of basis vectors. Further, suppose that the basis for $L^2(\mu)$ includes graduating functions $\{f_j\}_{j=1}^{p}$ that constitute a classical response surface model. Then the response function g has an exact representation as the convergent series:

$$g(x) = \sum_{j=1}^{p} \beta_j f_j(x) + \sum_{i=0}^{\infty} \theta_i g_i(x), \qquad (5.1)$$

where the functions $\{g_i\}_{i=0}^{\infty}$ complete the basis. Note that (5.1) could be written as a single infinite summation; we set off the first p terms to emphasize the way in which the Fourier series approach generalizes classical response surface models.

The generalized Fourier series (5.1) contains infinitely many parameters and cannot be used to model experimental data unless some assumptions are made about them. We do so in the form of prior distributions. Suppose that:

$$\beta \sim N(\beta_0, V_1) \qquad (5.2a)$$

$$\theta_i \sim N(0, m_i^2) \quad \text{independent.} \qquad (5.2b)$$

As with the previous models, it will often be of interest to assign a vague prior to β and this can be done by considering limiting forms as $V_1^{-1} \to 0$.

The rationale behind these prior assumptions is similar to that for the Bayesian models described in Section 3. The terms in the second summation in (5.1) can be thought of as the residual part of the response function that the classical model is unable to represent. Since the classical model is typically the best current

guess as to the nature of the response function, it seems reasonable to assume, a priori, that the $\{\theta_i\}$ will have zero means; their prior variances reflect the extent to which the experimenter is (or is not) confident that these terms make only a minimal contribution to the response function. For example, if the g_i are polynomials of increasing degree, then one might choose prior variances that decrease monotonically in i, progressively damping out the higher degree terms. A similar strategy might be invoked if the g_i are sines and cosines, with prior variances chosen to damp out the high frequency terms. The assumption that the $\{\theta_i\}$ are independent does not seem unreasonable provided we choose the basis functions to be an orthogonal sequence in $L^2(\mu)$.

5.2 Equivalence of the Fourier Approach and the Bayesian Models

We now prove that, under mild conditions, the generalized Fourier series approach is equivalent to the Bayesian models described in Section 3.

Theorem 5.1: Suppose an observed response variable $Y(x)$ is the sum of an unknown response function $g(x)$ and a random error:

$$Y(x) = g(x) + \varepsilon_x.$$

Suppose that the response function is modeled as a Bayesian generalized Fourier series:

$$g(x) = \sum_{j=1}^{p} \beta_j f_j(x) + \sum_{i=0}^{\infty} \theta_i g_i(x),$$

where, a priori,

$$\beta \sim N(\beta_0, V_1) \quad \text{and} \quad \theta_i \sim N(0, m_i^2),$$

subject to the restraint that $\sum_{i=0}^{\infty} m_i^2 g_i^2(x) < \infty$ for all $x \in X$.

Then $Y(x)$ follows the Blight-Ott model described in Section 3 (or equivalently the Smith model or the O'Hagan model). If, in addition, the explanatory variable space is compact, then the Bayesian models in Section 3 admit a generalized Fourier series representation of the above form.

Proof: First, define:

$$\eta_x = \sum_{i=0}^{\infty} \theta_i g_i(x). \tag{5.3}$$

Given the above prior specification for the $\{\theta_i\}$, and the restraint on the prior variances, η_x is a Gaussian process defined on the explanatory variable space X with:

$$E\{\eta_x\} = 0 \quad \forall\, x \in X, \tag{5.4a}$$

$$\text{Var}\{\eta_x\} = \sum_{i=0}^{\infty} m_i^2 g_i^2(x), \quad \text{and} \tag{5.4b}$$

$$\text{Cov}\{\eta_x, \eta_z\} = \sum_{i=0}^{\infty} m_i^2 g_i(x) g_i(z). \tag{5.4c}$$

The above distributional properties follow immediately from consideration of the limit of the characteristic function of the n'th partial sum.

The model for an observation at x can now be written:

$$Y(x) = \sum_{j=1}^{p} \beta_j f_j(x) + \eta_x + \epsilon_x,$$

where ϵ_x denotes the random error term for the observation. This is precisely the Bayesian model advocated by Blight and Ott (equation 3.3), with the approximating function given by the initial sum and the "bias" defined by η.

If the explanatory variable space is compact, it is also possible to deduce a Fourier series representation for any Blight-Ott model. First, suppose that there is only one explanatory variable. Gihman and Skorohod (1974) proved that any mean square continuous Gaussian process η_x defined on a closed interval of the real line admits the series expansion:

$$\eta_x = \sum_{i=0}^{\infty} \theta_i g_i(x). \qquad (5.5)$$

where the $\{\theta_i\}_{i=0}^{\infty}$ are independent, mean-zero, normal random variables. The expansion is derived by considering the covariance function of the Gaussian process as an integral operator (i.e. the kernel of an integral transform). It then follows from Mercer's Theorem (see Courant and Hilbert 1953, p. 138) that the covariance function can be expanded in terms of the eigenvalues and

eigenfunctions of the integral operator providing an analogue to equation (5.4c). Moreover, this series expansion converges absolutely and uniformly, not just in a normed sense. The series expansion of η_x then follows directly. The extension of Gihman and Skorohod's proof to compact domains in higher dimensional space is straight-forward, since the relevant theorems for integral equations are still valid (see, for example, Zabreyko, et. al. 1975, pp. 61-62).

The generalized Fourier series approach provides a potentially useful way to interpret the Bayesian models discussed earlier and emphasizes another way in which these models generalize classical response surface models, by adding extra regression functions whose coefficients are assumed, a priori, to be small.

5.3 The Significance of Vague Priors

Smith, Blight and Ott, and O'Hagan all advocated the use of a vague prior distribution for β, the vector of coefficients in the first summation of (5.1), and it was shown in Section 4 that a vague prior leads to generalized spline estimates. The special case of a vague prior for such parameters was also studied in detail by Lindley and Smith (1972). The generalized Fourier series representation of these models lends insight into the importance of assuming a vague prior. In particular, consider what would happen to the generalized Fourier series model (5.1) if a confirmed Bayesian statistician, with an aversion to vague priors, were to

assign proper prior distributions to all the coefficients i (5.1). The resulting model could then be written as a spec of (5.1) in which only the second summation appears. In Bl Ott's terminology, such a model would have no approximatinc function, only bias.

If, however, some of the coefficients in (5.1) are ass improper prior distributions, those terms cannot be include second summation without violating the restraint that the h finite prior variance at all points. Consequently, all ter coefficients have an improper prior must be treated separat all terms whose coefficients have a proper prior so that th division of (5.1) into two summations is not merely for cor rather, it is a necessary implication of the use of vague p the corresponding coefficients. The use of vague priors fc coefficients thus results in a fundamentally different mode

5.4 Relation to Ridge Regression

Hoerl and Kennard (1970) discussed the use of ridge re estimates for linear models $\mathbf{Y} = \mathbf{X}\beta + \epsilon$ in which there is multicollinearity of the columns of the $\mathbf{X}$ matrix. They c family of biased estimates of β by:

$$\tilde{\beta}(k) = (\mathbf{X}'\mathbf{X} + kI)^{-1}\mathbf{X}'\mathbf{Y},$$

where k is a parameter chosen to "stabilize" the ill-con matrix $\mathbf{X}'\mathbf{X}$. It is often recommended that the original re variables be centered and scaled before applying (5.6).

Hoerl and Kennard (1970) observed that the estimators (5.6) could also be given a Bayesian justification, as the posterior mean estimate given the prior assumption that

$$\beta \sim N(0, k^{-1}I), \tag{5.7}$$

that is, given the prior assumption that the regression coefficients are close to the origin. This fact has led some authors to criticize indiscriminate use of ridge regression without thought as to whether the above assumption is plausible (see, for example, Draper and Smith 1981, pp. 322-324 and the references therein).

Assumption (5.2b) of the generalized Fourier series approach is exactly analogous to (5.7), so that the Bayesian models can also be interpreted as complex ridge regression models. There are, however, several important differences. In the Bayesian models, the parameter of regression coefficients may be infinite and the prior covariance matrix is assumed to be diagonal but not proportional to the identity. The generalized Fourier series approach introduces assumption (5.2b) precisely because it is assumed to accurately reflect prior belief about the regression coefficients in the second summation of (5.1); the numerical considerations that inspired ridge regression play no role at all. Finally, the Fourier series approach applies this assumption to coefficients for the "extra" terms not included in a standard model; ridge regression applies the assumption to the original terms (the first summation in (5.1)) and does not add any extra terms.

5.5 Relation to Multiple Regression

Another special case of the generalized Fourier series is a Bayesian version of conventional multiple regression, in which the sampling theory model (2.3) is augmented with prior information about the regression coefficients, but no extra regression functions are introduced. If all p terms in the multiple regression model are assigned improper priors, the Bayesian model yields the ordinary least squares estimates of the parameters, so that ordinary least squares multiple regression is also a special case of (5.1). An alternative, and instructive, way to derive this special case is to delete the extra functions in (5.1) by requiring their coefficients to be 0; that is, assume that

$$g(x) = \sum_{i=0}^{\infty} \theta_i g_i(x),$$

where the single summation includes the p terms corresponding to the multiple regression functions, and assume that $\theta_i \sim N(0,m_i^2)$, where $m_i^2=\infty$ if g_i is one of the p multiple regression functions, and $m_i^2=0$ otherwise. Thus ordinary least squares multiple regression can be derived as a special case of (5.1) in which the variances of the model coefficients are allowed to take on only two values, 0 or ∞. The great flexibility of the Bayesian approach lies precisely in the ability to assign the prior variances intermediate values between these two extremes. As Morris (1983) observed, commenting on a similar model, a much richer class of

models is made available by considering positive, but finite, prior variances for model parameters.

5.6 Relation to Young's Method for Polynomial Regression

Young (1977) proposed a Bayesian method for polynomial regression in which a response to a single input variable is represented by an expansion in terms of a large, but finite, number of orthogonal polynomials, with prior distributions assigned to the coefficients of the polynomials. Clearly, such a polynomial expansion is a special case of the generalized Fourier series model (5.1). Young argued that "the best approach to prediction using polynomials is to fit the largest possible degree commensurate with our computing and statistical skills" (p. 309). As the results of this section make clear, there is no need to impose any maximal degree on the terms included in a Bayesian polynomial model, provided that the $\mathbf{X}$ matrix corresponding to those terms that are assigned improper prior distributions has full column rank and that the restraint cited in Theorem 5.1 is satisfied.

6. DEFINING PRIOR DISTRIBUTIONS

It is clear that the prior covariance structure is an important aspect of the Bayesian models discussed here. In particular, it can be shown that the form of estimates from these models depends on the form of the covariance function $V(x_1, x_2)$ (see Wahba 1978 and Steinberg 1983). Thus considerable thought must be devoted to

selecting a plausible covariance function. In this Section, we show how the generalized Fourier series approach discussed in Section 5 might be used to derive reasonable prior distributions.

Following Young's (1977) suggestion, a useful way to approximate a response which is a function of a single variable is to expand the response function in terms of a set of orthogonal polynomials: $\{P_i(x)\}_{i=0}^{\infty}$ where $P_i(x)$ is a polynomial of degree i. We will consider in detail the use of the Hermite polynomials, $\{H_i(x)\}_{i=0}^{\infty}$, to represent a response function in terms of a single input variable. The classical Hermite polynomials are defined to be orthogonal on the entire real line with respect to the measure space induced by the weight function $w(x) = \exp(-x^2)$. (See Szegö 1978, pp. 105-110 for basic properties of the classical Hermite polynomials.) We will assume, instead, that the weight function has been normalized to have measure 1, making it a normal (0,1/2) density function, and that the polynomials have been normalized to have square integral 1. Denoting the normalized Hermite polynomials by $H_i^*(x)$ and the classical polynomials by $H_i(x)$:

$$H_i^*(x) = 2^{-i/2}(i!)^{-1/2}H_i(x), \quad i=0,1,... \tag{6.1}$$

We consider expansions of the response function of the form:

$$g(x) = \sum_{i=0}^{\infty} \theta_i H_i^*(x). \tag{6.2}$$

Suppose, first, that proper priors are assigned to all the coefficients in (6.2); then the corresponding covariance function

will have the form:

$$V(x_1,x_2) = \sum_{i=0}^{\infty} m_i^2 H_i(x_1)H_i(x_2). \qquad (6.3)$$

For arbitrary choices of the prior variances, m_i^2, there does not appear to be a closed form solution for this series. A closed form solution does exist, however, for a useful parametric family: if $m_i^2 = \tau\sigma^2 w^i$, so that the prior variances decrease exponentially in the degree of the polynomial, then (6.3) is given by a slight modification (to account for the normalization) of Mehler's formula (see Watson 1933):

$$V(x_1,x_2;w) = \tau\sigma^2 \sum_{i=0}^{\infty} w^i H_i^*(x_1)H_i^*(x_2) \qquad (6.4)$$

$$= \tau\sigma^2(1-w^2)^{-1/2}\exp\{[2x_1x_2w - (x_1^2 + x_2^2)w^2]/(1-w^2)\}$$

$$= \tau\sigma^2(1-w^2)^{-1/2}\exp\{-(x_1-x_2)^2w^2/(1-w^2)\}\exp\{2wx_1x_2/(1-w)\},$$

where σ^2 measures the magnitude of experimental error, τ reflects the extent of the bias relative to experimental error, and $w \in [0,1)$ and controls the rate at which polynomials of increasing degree are discounted; small values of w correspond to prior belief that only polynomials of low degree are likely to be important components of the response function and large values of w reflect prior belief that higher-degree polynomials may also be important. The family of covariance functions defined by (6.4)

provides a flexible class of representations in which the parameters have straightforward interpretations in terms of prior beliefs about the nature of the experimental response.

It was assumed above that proper prior distributions would be assigned to all the coefficients in the expansion of the response function. It is a simple matter to modify (6.4) if it is desired to assign improper priors to coefficients for some of the low degree terms: one need only break the expansion into a finite summation, containing the terms whose coeeficients are assigned improper priors, and a second summation containing the rest of the terms, as in (5.1). The second summation would then correspond to the "bias" component of the model and its covariance function could be found by subtracting off the appropriate terms from (6.4). For example, if improper priors were assigned to the constant and linear terms in (6.2), the resulting covariance function would be:

$$V_1(x_1,x_2;w) = \tau\sigma^2 \sum_{i=2}^{\infty} w^i H_i^*(x_1)H_i^*(x_2) \qquad (6.5)$$

$$= V(x_1,x_2;w) - \tau\sigma^2\left[H_0^*(x_1)\ H_0^*(x_2) + wH_1^*(x_1)H_1^*(x_2)\right].$$

We should remark that the modification described in the preceding paragraph is actually unnecessary: if the model contains a linear and constant term whose coefficients are assigned improper priors, then using either (6.4) or (6.5) to define the bias results in the same model for the response function. On the surface, the last statement may seem surprising, since a model that contains the

same terms in the approximating model and in the bias would appear to suffer from problems of identifiability. To understand why the same model results in either case, consider for a moment just the constant term. Suppose the original covariance function (6.4) is used so that the approximating model contains the term $\beta_0 H_0^*(x)$ and the bias contains the term $\theta_0 H_0^*(x)$, where θ_0 has the proper prior distribution $\theta_0 \sim N(0,\tau\sigma^2)$, and β_0 is assigned an improper prior by assuming that $\beta_0 \sim N(0,k)$ and considering limiting forms as $k \to \infty$. These two terms, however, can be combined into the single term $\gamma_0 H_0^*(x)$, where $\gamma_0 = \beta_0 + \theta_0$. The prior distribution for γ_0 would then be: $\gamma_0 \sim N(0,\tau\sigma^2 + k)$. Considering limiting forms as $k \to \infty$ thus assigns an improper prior to γ_0. The resulting model is precisely that which would result had the modified covariance function (6.5) been used, with the single term $\gamma_0 H_0^*(x)$ in the approximating model. Of course, the parameters in the model do depend on which covariance function is used, since the single coefficient when (6.5) is used is the sum of the two coeffients that appear when (6.4) is used (it is here that the identifiability problem resides). In terms of the model for the response function, however, the two models are equivalent.

Although the discussion above has been restricted to the use of Hermite polynomials to represent the response function, other sets of orthogonal polynomials or, more generally, of orthogonal functions could also be used. For Jacobi polynomials, a formula analogous to (6.4) is given by Bailey (1938); however, it is more

complicated than (6.4) and requires the evaluation of a trigonometric integral to obtain the value of the covariance function for each pair of points.

One useful aspect of the Hermite polynomial expansion is that it has a natural extension to higher dimensions. Suppose $\mathbf{u}$ and $\mathbf{v}$ are vectors in k-dimensional Euclidean space and define:

$$V_k(\mathbf{u},\mathbf{v};w) = \tau\sigma^2 R_k(\mathbf{u},\mathbf{v};w), \quad \text{where} \tag{6.6}$$

$$R_k(\mathbf{u},\mathbf{v};w) = \exp\{-(\mathbf{u}-\mathbf{v})'(\mathbf{u}-\mathbf{v})w^2/(1-w^2)\} \times \exp\{2w\mathbf{u}'\mathbf{v}/(1-w)\} \; / \; (1-w^2)^{k/2}. \tag{6.7}$$

This clearly reduces to (6.4) when k=1, and it is easy to show that (6.6) is a legitimate covariance function on $R^k \times R^k$.

It is also possible to represent V_k by an expansion in terms of Hermite polynomials. For any $\mathbf{u}, \mathbf{v} \in R^k$, we can write:

$$R_k(u,v;w) = \prod_{j=1}^{k} R_1(u_j, v_j; w).$$

where the jth term in the product on the right-hand side of the equation is proportional to the one-dimensional covariance function (6.4) evaluated at the jth coordinates of $\mathbf{u}$ and $\mathbf{v}$. Substituting in the series expansion that led to the one-dimensional covariance function yields:

$$R_k(\mathbf{u},\mathbf{v};w) = \prod_{j=1}^{k} \sum_{i=0}^{\infty} w^i H_i^*(u_j) H_i^*(v_j), \tag{6.8}$$

a Cauchy product of series expansions in the normalized Hermite

polynomials for each of the k coordinates.

To interpret (6.8), it is easiest to consider first the case $k=2$. It is not difficult to rewrite the Cauchy product as:

$$R_2(\mathbf{u},\mathbf{v};w) = \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} w^{i+j} H_i^*(\mathbf{u}_1) H_i^*(\mathbf{v}_1) H_j^*(\mathbf{u}_2) H_j^*(\mathbf{v}_2). \quad (6.9)$$

Thus the covariance function for $k=2$ involves cross products of the terms in the one-dimensional (coordinatewise) covariance functions, with each term discounted exponentially in accord with the sum of the degrees of the respective polynomials. Moreover, (6.9) suggests (by analogy to (6.4)) how to define a stochastic process on R^2 that has the covariance function V_2. Let:

$$g(\mathbf{u}) = \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \theta_{i,j} H_i^*(\mathbf{u}_1) H_j^*(\mathbf{u}_2), \quad \text{for } \mathbf{u} \in R^2, \quad (6.10)$$

and suppose that the coefficients $\{\theta_{i,j}\}$ are independent normal random variables with mean 0 and variance $\tau\sigma^2 w^{i+j}$. Then $g(\mathbf{u})$ is a Gaussian stochastic process with covariance function V_2.

A natural interpretation of (6.10) is to view $g(\mathbf{u})$ as an expansion in terms of the two-dimensional orthogonal polynomials $\{H_i^*(\mathbf{u}_1) H_j^*(\mathbf{u}_2)\}_{i,j=0}^{\infty}$. It is thus also a natural Fourier series extension of classical two-dimensional response surface models which would begin by including linear terms in each coordinate, then add pure and mixed quadratic terms, then all cubic terms, etc, so that the degree d model would have the same form as (6.10), but with

the summation extending over all i and j for which $i + j \leq d$, rather than over all possible combinations. The notion that all terms of the same degree should be treated similarly is reflected above in the assumption that the prior variance of all coefficients of dth degree terms is equal to $\tau\sigma^2 w^d$.

For $k>2$, the Cauchy product (6.8) can be expanded in an exactly analogous manner. Again, the expansion can be shown to correspond to a generalization of classical k-variate response surface models. The exact formulas involve rather cumbersome notation and will be omitted.

The particular generalized Fourier series expansion used here, based on Hermite polynomials, seems intuitively appealing in the response surface context because it provides a natural generalization of the classical response surface models. It provides a simple parametric form for the covariance function and has the attractive property that it can be readily extended to handle several explanatory variables. It is, however, only one among many alternatives, and we suspect that consideration of other expansions will suggest additional useful covariance functions.

9. DISCUSSION

Scientists often use empirical models to describe the relationship between a response variable and a collection of explanatory variables. The models presented here all propose to account for the inability of any empirical graduating function to

perfectly represent an unknown response function. We have shown that the models of Smith (1973), Blight and Ott (1975), Young (1977), and O'Hagan (1978), although they are expressed in slightly different forms, are in fact equivalent to one another. We have also shown that they are closely related to smoothing splines (see Wahba 1978) and have argued that consideration of generalized Fourier series expansions of the response function provides a useful canonical form for the models.

We find it interesting that all the above authors were led to consider Bayesian models for the problem of representing model inadequacy. With conventional linear regression models, the only allowance made for the possibility that the given regression model may be inadequate is to posit a model that includes additional regression functions (e.g., if a straight line doesn't provide an adequate fit, use a quadratic). Of course, only a finite number of regression functions can be used and, as their number approaches the number of observations, the estimates can be quite unstable. The Bayesian approach seems to provide a much more realistic alternative for representing model inadequacy. By allowing the model to include an arbitrary number of regression functions, provided they are suitably downweighted, we can achieve a unified approach to model inadequacy instead of ad hoc attempts to find a combination of regression functions that fits the observed data. We think that Bayesian models such as those described here offer the only reasonable approach to the problem of model inadequacy.

References

Aronszajn, N. (1950). Theory of reproducing kernels. Trans. Amer. Math. Soc., 68, 337-404.

Bailey, W. N. (1938). The generating function of Jacobi polynomials. Jour. of the London Math. Soc., 13, 8-12.

Blight, B. J. N. and Ott, L. (1975). A Bayesian approach to model inadequacy for polynomial regression. Biometrika, 62, 79-88.

Box, G. E. P. (1954). The exploration and exploitation of response surfaces: Some general considerations and examples. Biometrics, 10, 16-60.

Box, G. E. P. and Wilson, K. B. (1951). On the experimental attainment of optimum conditions. Jour. Roy. Stat. Soc., B, 13, 1-45.

Box, G. E. P. and Youle, P. V. (1955). The exploration and exploitation of response surfaces: An example of the link between the fitted surface and the basic mechanism of the system. Biometrics, 11, 287-323.

Courant, R. and Hilbert, D. (1953). Methods of Mathematical Physics. New York: Interscience Publishers, Inc.

Draper, N. R. and Smith, H. (1981). Applied Regression Analysis, Second Edition. New York: John Wiley and Sons, Inc.

Gihman, I. I. and Skorohod, A. V. (1974). The Theory of Stochastic Processes I. New York: Springer-Verlag.

Hoerl, A. E. and Kennard, R. W. (1970). Ridge regression: biased estimation for non-orthogonal problems. Technometrics, 12, 55-67.

Lindley, D. V. and Smith, A. F. M. (1972). Bayes estimates for the linear model. Jour. Roy. Stat. Soc., B, 34, 1-41.

Morris, C. (1983). Parametric empirical Bayes inference: Theory and applications. Jour. Amer. Stat. Assoc., 78, 47-65.

Myers, R. H. (1976). Response Surface Methodology.

O'Hagan, A. (1978). Curve fitting and optimal design for prediction. Jour. Roy. Stat. Soc., B, 40, 1-41.

Rudin, W. (1974). Real and Complex Analysis, second edition. New York: McGraw-Hill Book Co.

Smith, A. F. M. (1973). Bayes estimates in one-way and two-way models. Biometrika, 60, 319-329.

Steinberg, D. M. (1983). Bayesian models for response surfaces and their implications for experimental design. Unpublished Ph. D. dissertation. University of Wisconsin-Madison.

Szegö, G. (1975). Orthogonal Polynomials, fourth edition. Providence: American Mathematical Society.

Wahba, G. (1978). Improper priors, spline smoothing and the problem of guarding against model errors in regression. Jour. Roy. Stat. Soc., B, 40, 364-372.

Watson, G. N. (1933). Notes on generating functions of polynomials: (2) Hermite polynomials. Jour. of the London Math. Soc., 8, 194-199.

Young, A. S. (1977). A Bayesian approach to prediction us
polynomials. Biometrika, 64, 309-317.

Zabreyko, P. P., Koshelev, A. I., Krasnosel'skii, M. A., M
G., Rakovshchik, L. S. and Stet'senko, V. Ya. (1975).
Equations - A Reference Text. trans. by T. O. Shaposh
S. Anderssen and S. G. Mikhlin, Leyden: Noordhoff Inte
Publishing.

SECURITY CLASSIFICATION OF THIS PAGE *(When Data Entered)*

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER 2682	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE *(and Subtitle)* BAYESIAN MODELS FOR RESPONSE SURFACES I: THE EQUIVALENCE OF SEVERAL MODELS AND THEIR RELATIONSHIP TO SMOOTHING SPLINES		5. TYPE OF REPORT & PERIOD COVERED Summary Report - no specific reporting period 6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) David M. Steinberg		8. CONTRACT OR GRANT NUMBER(s) DAAG29-80-C-0041 MCS-8210950
9. PERFORMING ORGANIZATION NAME AND ADDRESS Mathematics Research Center, University of Wisconsin 610 Walnut Street Madison, Wisconsin 53706		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS Work Unit Number 4 - Statistics & Probability
11. CONTROLLING OFFICE NAME AND ADDRESS See Item 18 below.		12. REPORT DATE April 1984 13. NUMBER OF PAGES 38
14. MONITORING AGENCY NAME & ADDRESS*(if different from Controlling Office)*		15. SECURITY CLASS. *(of this report)* UNCLASSIFIED 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT *(of this Report)*

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT *(of the abstract entered in Block 20, if different from Report)*

18. SUPPLEMENTARY NOTES

U. S. Army Research Office
P. O. Box 12211
Research Triangle Park
North Carolina 27709

National Science Foundation
Washington, D. C. 20550

19. KEY WORDS *(Continue on reverse side if necessary and identify by block number)*

Response Surface Models
Bayesian Linear Model
Hierarchical Linear Model
Localized Regression Model
Smoothing Splines
Polynomial Regression
Series Estimators

20. ABSTRACT *(Continue on reverse side if necessary and identify by block number)*

In response surface modeling, simple graduating functions such as low-degree polynomials are used to approximate complex, unknown response functions. Several authors have suggested Bayesian generalizations of response surface models that incorporate prior belief as to the (in)adequacy of a graduating function to represent a response function. We show that the models of Smith (1973), Blight and Ott (1975), and O'Hagan (1978) are equivalent. We also show

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE *(When Data Entered)*

20. ABSTRACT - cont'd.

how their models are related to the generalized smoothing splines of Wahba (1978) and to Young's (1977) proposal for Bayesian polynomial regression. Finally, we suggest a canonical representation of the models in terms of generalized Fourier series expansions of the response function and show how such expansions can be used to develop reasonable prior distributions.